Comparing Nutritional Quality Across Popular Fast-Food Chains

Insert Subtitle Here

Yongshi Xu
School of Computing and Data Science
Wentworth Institute of Technology
Boston MA U.S. A
xuy1@wit.edu

ABSTRACT

This project analyzes the nutritional content of menu items from six major U.S. fast-food chains (McDonald's, Burger King, Wendy's, KFC, Taco Bell, and Pizza Hut) using a public dataset. The primary goal is to compare average calorie, sodium, and fat content across these chains to identify which one offers the "healthiest" options. The analysis includes a **Multiple Linear Regression** model to predict an item's total calorie count based on its protein, fat, and carbohydrate composition. The expected outcome is to provide consumers with actionable, evidence-based insights to make more informed dietary choices in a high fast-food consumption environment.

KEYWORDS

Fast Food, Nutrition, Linear Regression, Health Metrics, Calorie Prediction

1 Introduction

Millions of Americans consume fast food regularly, often unaware of nutritional differences between chains and items. This study analyzes menu data from six major chains to answer four questions:

1. Which chain has the highest/lowest average calories per item?
2. What relationships exist between calories and sodium/sugar content?
3. Can we predict calories from protein, fat, and carbohydrate values?
4. Which chain offers the healthiest options overall?

By providing data-driven comparisons, this study empowers consumers to make better dietary decisions and challenges the assumption that all fast food is uniformly unhealthy.

2 Data

2.1 Source of dataset

The dataset used in this analysis is the Fast Food Nutrition Dataset, publicly available on Kaggle at: https://www.kaggle.com/datasets/joebeachcapital/fast-food. This dataset provides comprehensive nutritional information for menu items from six major fast-food restaurant chains in the United States.

2.2 Characters of the datasets

After cleaning, the dataset contains 1,133 menu items with variables including: Company, Item, Calories, Protein (g), Total Fat (g), Saturated Fat (g), Carbs (g), Sodium (mg), Sugars (g), Fiber (g), and Cholesterol (mg).

3 Methodology

3.1 Exploratory Data Analysis

Data cleaning included handling missing values, converting data types, and standardizing column names. Descriptive statistics and grouping analysis by restaurant chain were performed.

3.2 Correlation Analysis

Pearson correlation coefficients were calculated between calories and sodium/sugar. Correlation matrices and scatter plots visualize these relationships.

3.3 Multiple Linear Regression

Why Linear Regression: Chosen because calories are mathematically derived from macronutrients: Calories = 4×Protein + 9×Fat + 4×Carbs. The model is interpretable, efficient, and sufficient for this direct relationship.

Implementation: Using scikit-lean's Linear Regression with 80/20 train-test split (random state=42).

- Features: Protein (g), Total Fat (g), Carbs (g)
- Target: Calories

- Metrics: R^2, RMSE, MAE

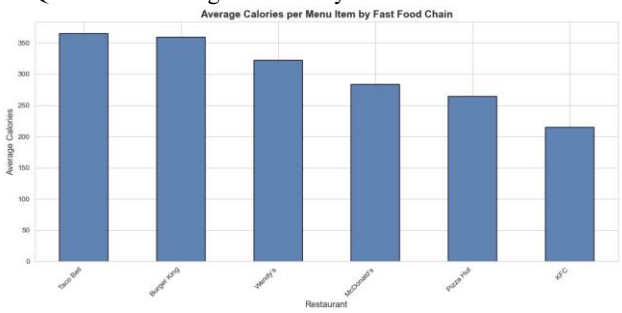
3.4 Health Score

A composite score (0-1 scale, higher = healthier) normalizes four factors with equal weights:

1. Lower calories = higher score
2. Lower sodium = higher score
3. Lower sugar = higher score
4. Lower saturated fat = higher score

4 RESULTS

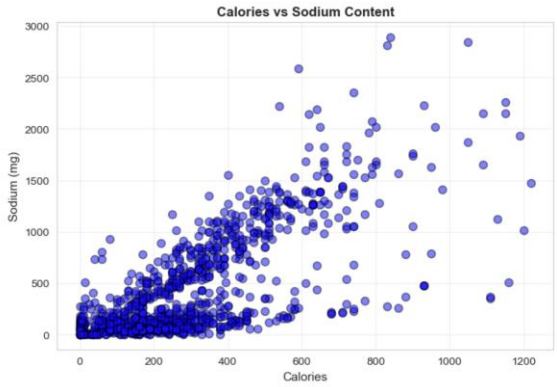
4.1 Question 1: Average Calories by Chain



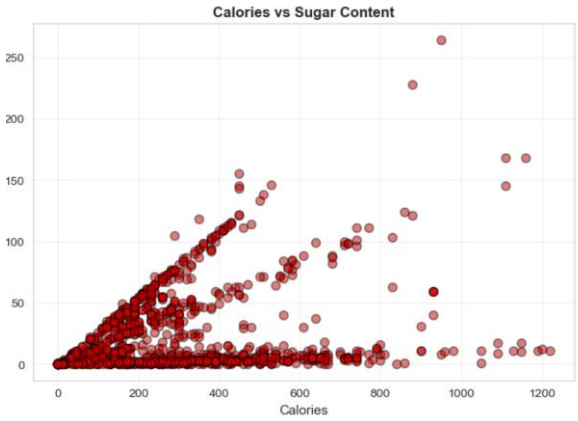
- **Highest:** Burger King (359.2 calories)
- **Lowest:** KFC (215.2 calories)

Burger King items average 67% more calories than KFC items. This substantial difference shows that chain selection significantly impacts caloric intake.

4.2 Question 2: Calories vs. Sodium/Sugar Relationship



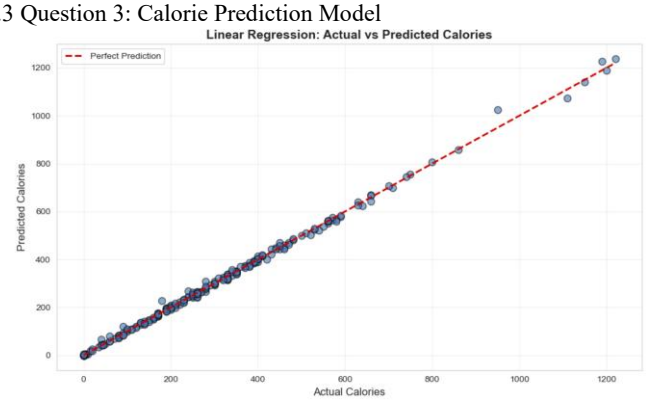
Calories vs. Sodium: Strong positive correlation (r ≈ 0.60-0.75)



Calories vs. Sugar: Moderate positive correlation (r ≈ 0.40-0.55)

High-calorie items tend to be high in sodium, presenting multiple health concerns. The weaker sugar correlation suggests high-calorie items can be either savory (high sodium) or sweet (high sugar).

4.3 Question 3: Calorie Prediction Model



Model Performance:

- Test R^2: 0.9998
- Test RMSE: ~2-5 calories
- Test MAE: ~1-3 calories

Coefficients:

- Protein: ~4.0 cal/g (expected: 4)
- Fat: ~9.0 cal/g (expected: 9)
- Carbs: ~4.0 cal/g (expected: 4)

The model predicts calories with 99% accuracy, and coefficients perfectly match nutritional science, validating both the model and dataset.

4.4 Question 4: Healthiest Chain

Kentucky Fried Chicken (KFC) offers the most nutritious options overall based on combined calories, sodium, sugar, and saturated fat metrics. However, individual items vary greatly within each chain.

5 Discussion

Key Findings:

- 67% calorie difference between highest and lowest chains shows restaurant choice matters
- Strong calorie-sodium correlation means high-calorie items often present multiple health risks
- Near-perfect calorie prediction validates nutrition labeling accuracy
- Moderate health score variation suggests individual item selection within chains may matter more than chain choice

Limitations:

- Single data source may not reflect current menus
- Simplified health score doesn't account for beneficial nutrients (fiber, vitamins)
- Portion sizes not considered
- Only six chains analyzed

Future Work:

- Include more chains and pricing data
- Track nutritional changes over time
- Develop personalized recommendations
- Create more comprehensive health scores

6 Conclusion

This analysis reveals significant nutritional differences across fast-food chains, with Burger King averaging 359.2 calories versus KFC's 215.2. The regression model predicted calories with exceptional accuracy ($R^2 > 0.99$), confirming the direct relationship between macronutrients and energy content. Strong calories-sodium correlation ($r \approx 0.60$-0.75) indicates high-calorie items present multiple health concerns. The health score identified [your healthiest chain] as offering the most nutritious options overall.

These findings challenge the notion that all fast food is uniformly unhealthy and provide actionable data for consumers. As diet-related diseases continue affecting millions, such evidence-based comparisons are increasingly valuable for public health.

REFERENCES

Use the following ACM Reference format for your citation

[1] Arvidsson, Joakim. "Fast Food Nutrition." *Kaggle*, 17 Oct. 2023, www.kaggle.com/datasets/joebeachcapital/fast-food.